

DALEX: Descriptive mAchine Learning EXplanations. Tools for exploration, validation and explanation of complex machine learning models (M. Staniak, Warsaw University of Technology)

Interpretable Machine Learning is a new field of study which aims to overcome limitations in interpretability of black box models that hinder the development of Machine Learning methods in regulated areas such as finance and health sciences by providing method of explaining and validating complex statistical models.

In this talk I will provide a brief introduction to Interpretable Machine Learning as a new area of research and present R package DALEX which delivers a unified interface to different model explainers and an associated family of packages such as auditor, live and ceterisParibus.